IN THE SPECIFICATION:

(1)   On page 1, immediately under the heading "Title of the Invention", please delete the title, i.e., the paragraph starting on line 3 and ending on line 5, and replace it with the following new title:

Singing Voice Synthesis Apparatus and Method with Deterministic and Stochastic Components

(2)   On page 5, please replace the paragraph starting on line 29 and ending on line 34 with the following paragraph, amended as indicated therein:

It is a third object of the present invention to provide a singing voice synthesizing apparatus ~~apparatus~~ and a singing voice synthesizing method that are capable of adjusting the degree of huskiness in a synthesized voice, and a program for realizing a singing voice synthesizing method.

(3)   On page 8, please replace the paragraph starting on line 2 and ending on line 7 with the following paragraph, amended as indicated therein:

Still more preferably, when repeating the plurality of frames of the frame string corresponding to the data of the stochastic ~~compoenent~~ component of each of the voice fragments in the first and second directions, the duration time adjusting device reverses a phase of a phase spectrum of the stochastic component.

(4)   On page 12, please replace the paragraph starting on line 23 and ending on line 26 with the following paragraph, amended as indicated therein:

FIGS. 14A and 14B is a diagram illustrating a singing voice synthesis process carried out by the singing voice synthesizing apparatus according to the other embodiment of [[of]] the present invention;

(5)   On page 18, please replace the paragraph starting on line 7 and ending on line 14 with the following paragraph, amended as indicated therein:

Reference numeral 22 designates a deterministic component adjusting means that, based on control parameters such as pitch, dynamics and tempo that are included in the melody data of the song, adjusts the data of the deterministic component of fragment data read from the phoneme database 10, and reference numeral 23 ~~deisgnates~~ designates a stochastic component adjusting means that adjusts the data of the stochastic component.

(6)   Please replace the paragraph starting on line 15, page 18, and ending on line 2, page 19, with the following paragraph, amended as indicated therein:

Reference numeral 24 designates a duration time adjusting means that varies the duration time of fragment data output from the deterministic component adjusting means 22 and from the stochastic component adjusting means 23. Reference numeral 25 designates a fragment level adjusting means that adjusts the level of each fragment data output from the duration time adjusting means 24. Reference numeral 26 designates a fragment concatenating means that concatenates individual fragment data, which have been level-adjusted by the fragment level adjusting means 25, into a time series. Reference numeral 27 ~~desinates~~ designates a deterministic component generating means that, based on the deterministic components of fragment data that have been concatenated by the fragment concatenating means 26, generates

3

deterministic components (harmonic components) having a desired pitch. Reference numeral 28 designates an adding means that synthesizes harmonic components generated by the deterministic component generating means 27 and stochastic components output from the fragment concatenating means 26. Voice synthesis can be achieved by transforming the output from this adding means 28 into a time domain signal.

(7)   On page 20, please replace the paragraph starting on line 18 and ending on line 19 with the following paragraph, amended as indicated therein:

An ~~unvocied~~ unvoiced sound contains no deterministic component.

(8)   Please replace the paragraph starting on line 32, page 20, and ending on line 17, page 21, with the following paragraph, amended as indicated therein:

FIG. 3A is an example of an amplitude spectrum of a stochastic component obtained from an SMS analysis of a voiced sound. It is difficult to completely remove the effect of the deterministic component, and as shown in the figure, there are some peaks in the vicinity of the harmonics. If this stochastic component is used as it is, to synthesize a voice sound at a pitch different from the original pitch, peaks will appear in the vicinity of lower frequency harmonics, which do not ~~blendsmoothly~~ blend smoothly with the deterministic component and audible as a harsh sound. To avoid this, the frequency of the stochastic component may be varied so as to match a change in pitch. However, since high frequency stochastic components are less affected by the deterministic component, it is desirable to use the original amplitude spectrum as it is. In other words, in the low frequency region, it should be sufficient to compress and expand the frequency axis according to the desired pitch. However, the original tone color must not be

4

changed at this time. Namely, it is necessary that the general shape of the amplitude spectrum be preserved while carrying out this processing.

(9)   On page 28, please replace the paragraph starting on line 2 and ending on line 27 with the following paragraph, amended as indicated therein:

In FIGS. 9A and 9B, reference numeraal numeral 31 designates a lyric-melody separating means that separates lyric data and melody data from the music score data of a song for which a singing voice is to be synthesized, and 32 a lyric-to-phonetic code conversion means that converts the lyric data from the lyric-melody separating means 31 into a string of phonetically coded data (phonemes). A phoneme string from the lyric-to-phonetic code conversion means 32is means 32 is input to the phoneme (phonetic code)-to-fragment conversion means 21. Various control parameters, such as tempo, may be input to control the musical performance. Pitch information and dynamics information such as dynamic marks that has been separated from the music score data by the lyric-melody separating means 31, and the control parameters are input to a pitch determining means 33, which in turn determines the pitch, dynamics, and tempo of the signing sound. Fragment information from the phoneme-to-fragment conversion means 21 and information such as pitch, dynamics, and tempo from the pitch determining means 33 are fed to a fragment selecting means 34. The fragment selecting means 34 searches the voice fragment database (phoneme database) 10 and outputs the most suitable fragment data. At this time, if there is stored no fragment data that completely matches the search conditions, data of one or a plurality of similar fragments is read out.

5

20558946v1